

Quantifying Heteroskedasticity via Bhattacharyya Distance

M. Hassan, M. Hossny, D. Creighton and S. Nahavandi

September 21, 2016

Abstract

Heteroskedasticity is a statistical anomaly that describes differing variances of error terms in a time series dataset. The presence of heteroskedasticity in data imposes serious challenges for forecasting models and many statistical tests are not valid in the presence of heteroskedasticity. Heteroskedasticity of the data affects the relation between the predictor variable and the outcome, which leads to false positive and false negative decisions in the hypothesis testing. Available approaches to study heteroskedasticity thus far adopt the strategy of accommodating heteroskedasticity in the time series and consider it an inevitable source of noise. In these existing approaches, two forecasting models are prepared for normal and heteroskedastic scenarios and a statistical test is to determine whether or not the data is heteroskedastic.

This work-in-progress research introduces a quantifying measurement for heteroskedasticity. The idea behind the proposed metric is the fact that a heteroskedastic time series features a uniformly distributed local variances. The proposed measurement is obtained by calculating the local variances using linear time invariant filters. A probability density function of the calculated local variances is then derived and compared to a uniform distribution of theoretical ultimate heteroskedasticity using statistical divergence measurements. The results demonstrated on synthetic datasets shows a strong correlation between the proposed metric and number of variances locally estimated in a heteroskedastic time series.

1 Introduction

Quantifying heteroskedasticity is a relatively new approach to study this statistical artefact. While heteroskedasticity is dealt with as an inevitable source of noise that must be accounted for in forecasting models, it becomes a noise source in signal processing and machine learning techniques [1]. This new uncontrollable source of noise then creates a new challenge to quantify heteroskedasticity. Consequently, early solutions for quantifying heteroskedasticity adopted one of two schools, change point detection and local parameter estimation. Change

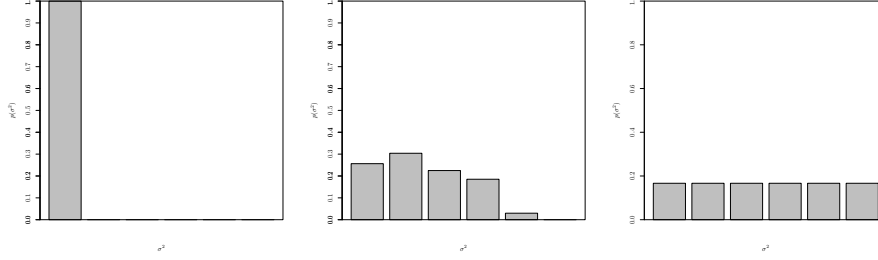


Figure 1: Probability density function of local variances for a homoskedastic [left] and a heteroskedastic [middle] and a theoretically ultimate heteroskedastic [right] time series.

point detection methods [2] utilise the available heteroskedasticity tests to perform a binary segmentation of a heteroskedastic time series into smaller homoskedastic fragments.

Local parameter estimation methods utilise convolution with linear time invariant filters to estimate local variance at every sample based on its neighbours within a certain window w . There are two methods that adopt the local parameter estimation approach. Heteroskedasticity Variance Index (HVI) derived a variance of local variance as an indication of heteroskedasticity [3]. Slope of Local Variance index (SoLVi) used the slope of the trend of estimated local variance derived by HVI as an indication of heteroskedasticity [4].

An alternative approach to measure heteroskedasticity is to sample the estimated local variances in the time series. By doing this, a probability distribution p_{σ^2} of the local variances can be derived. In theory, and as demonstrated in Figure 1, a homoskedastic time series should have a consistent local variance σ^2 over time. Consequently, the probability distribution of a homoskedastic time series should be unimodal and centred around σ^2 . On the other hand, a heteroskedastic time series should, in theory, approach a uniform distribution covering a wide range of local variances. The ultimate heteroskedasticity time series should, in theory, feature a uniform distribution $\mathcal{U}(0, \infty)$. Therefore, measuring the distance between the probability distribution of the local variances p_{σ^2} and the uniform distribution provides a quantified measure of heteroskedasticity. In this section, we propose heteroskedasticity measures based on probability distribution metrics. The heteroskedasticity quantified measure for a time series y is defined as follows:

$$\mathcal{H}(y) = \Delta_p(P(\sigma_y^2), \mathcal{U}(0, \infty)), \quad (1)$$

where $\Delta_p : \mathbb{P}^2 \rightarrow [0, 1]$ is a distribution distance function of the estimated local variances σ_y^2 . Many probability distribution metrics are available. However, most of these metrics rely on entropies, joint probability density functions and sigma algebra. In this work-in-progress research, a justification for exclud-

ing three of the most famous probability distribution metrics such as mutual information, Tsallis mutual information, and Jensen-Shannon divergence are discussed. Additionally, a heteroskedasticity quantification function based on Bhattacharyya distance is introduced.

The rest of this paper is organised as follows. Section 2 covers mutual information and its variations. Section 3 covers probability distribution divergence metrics based on Renyi entropy and introduces to the Bhattacharyya distance. Section 4 covers Bhattacharyya distance and the way it is utilised to quantify heteroskedasticity. Finally, Section 5 presents conclusion.

2 Mutual Information (MI)

Mutual Information between two random variables X and Y derives a cross entropy between the joint probability distribution $p(x, y)$ and the ultimate scenario of complete mutual independence $p(x) \cdot p(y)$ as follows:

$$MI(X; Y) = \iint p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) dx dy. \quad (2)$$

MI is used to measure the information shared between X and Y and equals to zero when X and Y are completely independent as follows:

$$\begin{aligned} MI(X; Y) &= \iint p(x)p(y) \log \left(\frac{p(x)p(y)}{p(x)p(y)} \right) dx dy \\ &= \iint p(x)p(y) \log 1 dx dy = 0. \end{aligned} \quad (3)$$

Mutual Information does not, however, provide a good solutions for quantifying heteroskedasticity because it is only bounded with the maximum entropy of X or Y as follows.

$$\begin{aligned} MI(X; X) &= \int p(x) \log \left(\frac{p(x)}{p(x)p(x)} \right) dx \\ &= \int p(x) \log \left(\frac{1}{p(x)} \right) dx \\ &= - \int p(x) \log p(x) dx = H(X), \end{aligned} \quad (4)$$

where $H(X)$ is the entropy of X .

2.1 Tsallis Driven Mutual Information (MI_α)

Another variation of mutual information was proposed by Cvejic *et al.* in [5]. They proposed to use the tunable Tsallis entropy [6] described below.

$$MI_\alpha(X, Y) = \frac{1}{1-\alpha} \left(1 - \iint \frac{p(x)^\alpha}{p(y)^{1-\alpha}} dx dy \right), \quad (5)$$

where $\alpha \in \mathbb{R} - \{1\}$ and $MI_\alpha(X, Y) \rightarrow MI(X, Y)$ as $\alpha \rightarrow 1$. This was proven by applying l'hopital rule on eq. 5 and substituting $\alpha = 1$.

$$\begin{aligned} MI_1(X, Y) &= \lim_{\alpha \rightarrow 1} MI_\alpha(X, Y) \\ &= \lim_{\alpha \rightarrow 1} \frac{\frac{d}{d\alpha} [1 - \iint p(x)^\alpha p(y)^{\alpha-1} dx dy]}{\frac{d}{d\alpha} [1 - \alpha]} \\ &= \lim_{\alpha \rightarrow 1} \iint p(x)^\alpha p(y)^{1-\alpha} \frac{\ln p(x)}{\ln p(y)} dx dy \\ &= MI(X, Y). \end{aligned} \quad (6)$$

2.2 Jensen-Shannon Divergence

Jensen-Shannon divergence metric uses sigma algebra [7] to derive an intermediate random variable $M = \frac{1}{2}(X+Y)$ which serves as a reference point to measure distance of X and Y from using mutual information as follows:

$$JSD(X, Y) = \frac{1}{2}MI(X, M) + \frac{1}{2}MI(Y, M). \quad (7)$$

While this metric is bounded to $0 \leq JSD(X, Y) \leq 1$, deriving the mixture distribution of the random variable M is computationally intensive.

3 Renyi Divergence

Renyi divergence [8] uses a generalised form of Shannon, Hartley, min-, and collision- entropies [7, 9] and is formulated as follows:

$$H_\alpha = \frac{1}{1-\alpha} \log \left(\int p(x)^\alpha \right). \quad (8)$$

Renyi's divergence metric is then formulated as follows.

$$\mathcal{R}_\alpha(X, Y) = \frac{1}{1-\alpha} \log \left(\iint p(x)^\alpha p(y)^{1-\alpha} dx dy \right). \quad (9)$$

As Renyi entropy generalises many entropies, its divergence metric also generalises many divergence metrics. For example, when $\alpha \rightarrow 1$ the Renyi entropy converges to Shannon's entropy and the divergence metric converges to the Mutual Information metric by applying l'hopital rule as follows:

$$\begin{aligned}
H_1(X) &= \lim_{\alpha \rightarrow 1} \frac{\log \left(\int p(x)^\alpha \right)}{1 - \alpha} \\
&= \lim_{\alpha \rightarrow 1} - \frac{\int p(x)^\alpha \log p(x)}{\int p(x)^\alpha} \\
&= - \frac{1}{\int p(x)} \int p(x) \log p(x) \\
&= - \int p(x) \log p(x) = H(X)
\end{aligned} \tag{10}$$

$$\begin{aligned}
\mathcal{R}_1 &= \iint p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) dx dy \\
&= MI(X, Y)
\end{aligned} \tag{11}$$

Additionally, Renyi divergence also correlates with Bhattacharyya coefficient when $\alpha = \frac{1}{2}$ as follows:

$$\begin{aligned}
\mathcal{R}_{\frac{1}{2}}(X) &= \frac{\log \left(\iint p(x)^{\frac{1}{2}} p(y)^{1-\frac{1}{2}} dx dy \right)}{1 - \frac{1}{2}} \\
&= -2 \log \left(\iint \sqrt{p(x)p(y)} dx dy \right) \\
&= -2 \log BC(X, Y) \\
&= 2\Delta_p^B(X, Y).
\end{aligned} \tag{12}$$

4 Bhattacharyya Heteroskedasticity Measure

4.1 Bhattacharyya Distance

Bhattacharyya-based metrics rely on deriving the Bhattacharyya Coefficient (BC) [10]. The BC measures the closeness between two probability distributions p and q by measuring how disjoint they are as follows:

$$BC(p, q) = \sum_{x \in X} \sqrt{p(x)q(x)} \tag{13}$$

Figure 2 shows the Bhattacharyya coefficient as number of local variances increase in the dataset. Bhattacharyya coefficient has an upper bound of 1 if and only if $p(x) = q(x)$.

This coefficient is then used to derive the Bhattacharyya distance as

$$\Delta_p^B(p, q) = -\ln BC(p, q) \tag{14}$$

However, this distance function has no upper bound and does not satisfy the triangulation inequality. Figure 3 demonstrates the Bhattacharyya distance.

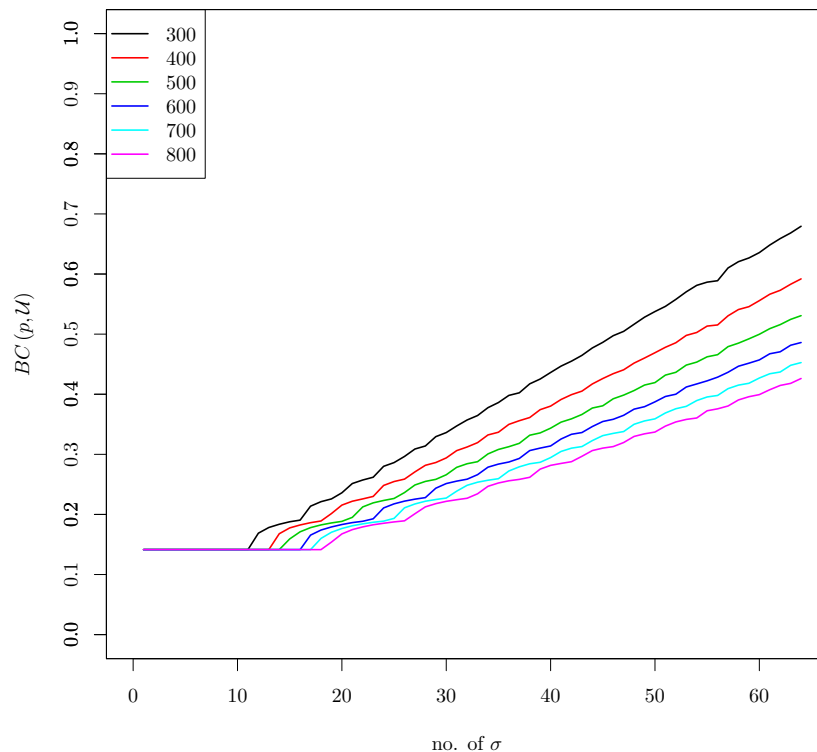


Figure 2: Bhattacharyya coefficient of a for time series generated using different 64 sigmas. The graphs demonstrate different kernel sizes w .

4.2 Hellinger Distance

Finally, Hellinger *et al.* provided a sound Bhattacharyya based divergence metric that is bounded and satisfies the triangulation inequality in [11]. The Hellinger metric is derived from Bhattacharyya coefficient as:

$$\Delta_p^H(p, q) = 1 - \sqrt{1 - BC(p, q)} \quad (15)$$

Figure 4 shows the effect of window size on Hellinger divergence metric.

As a heteroskedastic time series, by definition, is derived from systems of different variances; the probability distribution of local variances $p(\sigma)$ of a heteroskedastic time series must be approaching a uniform distribution \mathcal{U} . On the other hand a homoskedastic time series will have a probability distribution further from the uniform distribution \mathcal{U} . To guarantee bounded function we

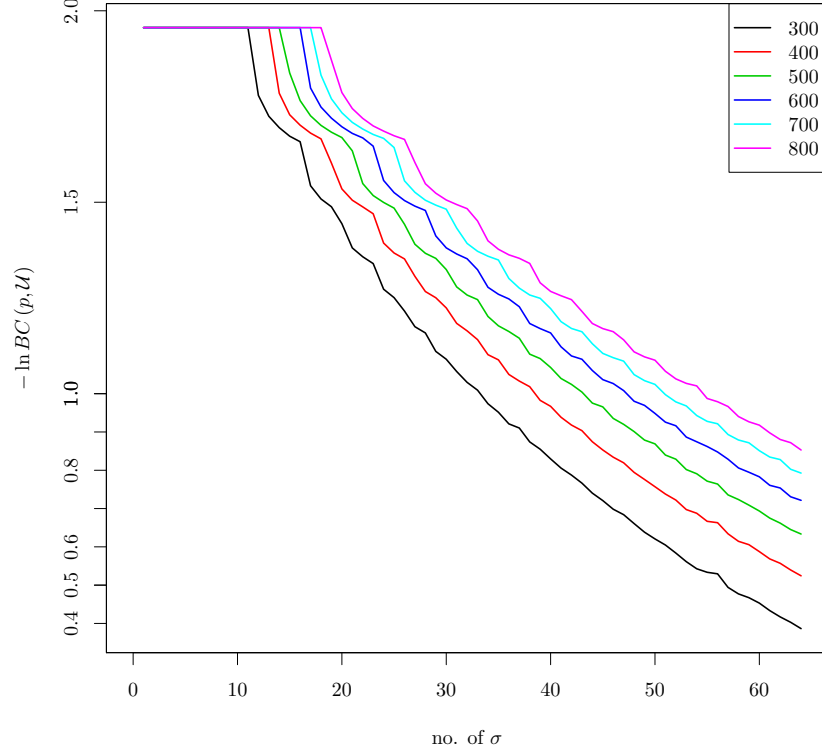


Figure 3: Bhattacharyya distance of a for time series generated using different 64 sigmas. The graphs demonstrate different kernel sizes w .

chose Bhattacharyya coefficient over the Renyi driven metric in eq. 9. The Bhattacharyya heteroskedasticity measure is then formulated as follows:

$$\mathcal{H}_B(y) = \sum_{x \in X} \sqrt{P(\sigma_y^2) \mathcal{U}(\sigma_y^2)} \quad (16)$$

where $P(\sigma_y^2)$ is a probability distribution function of the estimated local variances σ_y^2 . A Hellinger variation can also be derived with the same concept as follows:

$$\mathcal{H}_H(y) = 1 - \sqrt{1 - \sum_{x \in X} \sqrt{P(\sigma_y^2) \mathcal{U}(\sigma_y^2)}} \quad (17)$$

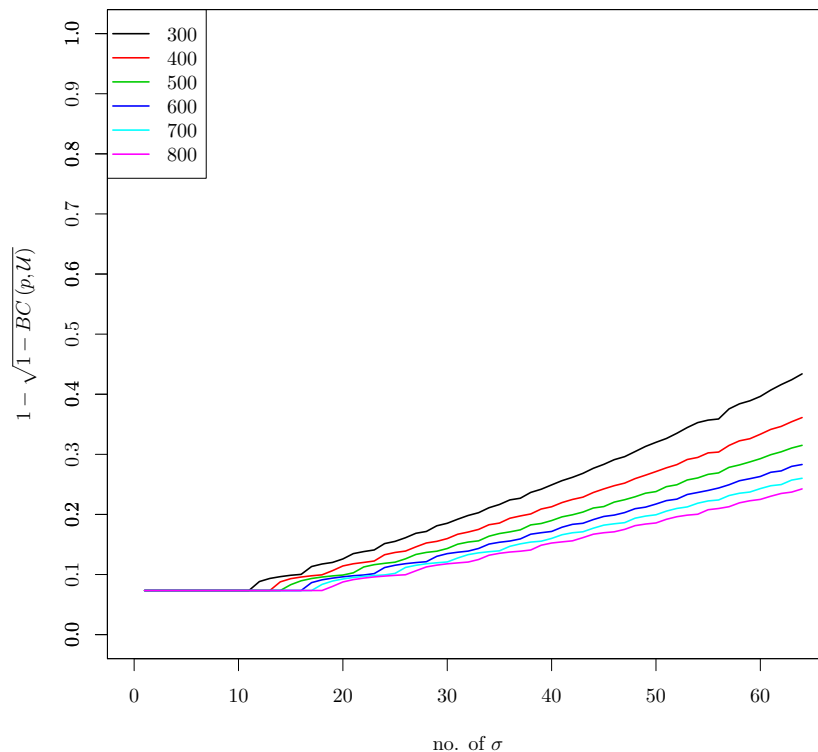


Figure 4: Hellinger coefficient of a for time series generated using different 64 sigmas. The graphs demonstrate different kernel sizes w .

5 Conclusions

In this paper, we examine the divergence heteroskedasticity measures. Our motivation is that most of the available probability distribution metrics rely on entropies, joint density functions and sigma algebra. Measuring the distance between the probability distribution of the local variances p_{σ^2} and the uniform distribution (ultimate heteroskedasticity) provides a quantified measure of heteroskedasticity. Consequently, the Bhattacharyya distance was adopted to introduce the Bhattacharyya heteroskedasticity measure. The main reason behind preferring the Bhattacharyya over the other KL-divergence measures is to guarantee a bounded function. The Bhattacharyya heteroskedasticity measure is then formulated using Hellinger variation to maintain the three properties of a distance function.

Acknowledgement

This research was fully supported by the Institute for Intelligent Systems Research and Innovation (IISRI).

References

- [1] A. Foi, “Clipped noisy images: Heteroskedastic modeling and practical denoising,” *Signal Processing*, vol. 89, no. 12, pp. 2609–2629, Dec. 2009.
- [2] M. Hassan, M. Hossny, S. Nahavandi, and D. Creighton, “Quantifying heteroskedasticity via binary decomposition,” *IEEE International Conference on Computer Modelling and Simulation*, pp. 112–116, 2013.
- [3] —, “Heteroskedasticity variance index,” *IEEE International Conference on Computer Modelling and Simulation*, pp. 135–141, 2012.
- [4] —, “Quantifying heteroskedasticity using slope of local variances index,” *IEEE International Conference on Computer Modelling and Simulation*, 2013.
- [5] N. Cvejic, C. Canagarajah, and D. Bull, “Image fusion metric based on mutual information and tsallis entropy,” *Electronics Letters*, vol. 42, no. 11, pp. 626–627, 2006.
- [6] C. Tsallis, “Possible generalization of boltzmann-gibbs statistics,” *Journal of Statistical Physics*, vol. 52, pp. 479–487, 1988.
- [7] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, no. 3, p. 379423, 1948.
- [8] A. Rnyi, “On measures of information and entropy,” *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability*, pp. 547–561, 1960.
- [9] R. Knig, R. Renner, and C. Schaffner, “The operational meaning of min- and max-entropy,” *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 4337–4347, 2009.
- [10] A. Bhattacharyya, “On a measure of divergence between two statistical populations defined by their probability distributions,” *Bulletin of the Calcutta Mathematical Society*, vol. 35, pp. 99–109, 1943.
- [11] E. Hellinger, “Neue begrndung der theorie quadratischer formen von unendlichvielen vernderlichen,” *Journal fr die reine und angewandte Mathematik*, vol. 136, pp. 210–271, 1909.